Incremental Reconstruction Approach for Telepresence or AR Applications

Luis Almeida
ISR, Univ. Coimbra
Polytechnic of Tomar
Tomar, Portugal
laa@ipt.pt

Paulo Menezes
ISR, Univ. Coimbra
Coimbra
Portugal
paulo@isr.uc.pt

Jorge Dias
ISR, Univ. Coimbra
Coimbra
Portugal
jorge@deec.uc.pt

Abstract

This paper proposes an on-line incremental 3D reconstruction framework aimed at fulfilling the needs of telepresence or human machine interaction applications. The research presents a teleconference system that improves and induces the feeling that persons are in the presence of each other. A free viewpoint method, based on realistic user's appearances, is proposed to simulate a real face-to-face meeting. The contributions are: a new incremental version of Crust algorithm that enables incremental fusion of sensor data and a confidence-based method that automatically decides whether or not to integrate newly acquired data in the existing model based on measure uncertainty and novelty. To avoid the classical stereo vision reconstruction problems, the method bases on hybrid sensors to acquire simultaneous depth information and the corresponding texture image (e.g. kinect). This enables the alignment between acquired data and pre-adquired model by maximizing a criterion that is related with the matching between visual features and between acquired shapes. A mesh based representation enables the use of the surface topological geometric information during the data model integration process.

Keywords

Three-Dimensional Graphics and Realism, Augmented Reality, Reconstruction, Range Data, Tracking, Telepresence

1 INTRODUCTION

Widely used video teleconference applications (ex: Cisco WebEx, Citrix GoToMeeting, Microsoft Skype, Google Hangouts or Apple Facetime) are not replicating important real face-to-face meeting cues, like eye-to-eye contact establishment, gesture reconnaissance, body language or facial expressions. Nevertheless, recent advances on sensing, display and computation technology are creating the ideal condition for affordable consumer 3D applications in Augmented Reality (AR), Virtual Reality (VR) or Human Machine Interactions (HMI). Our application concept goal is depicted in Figure 1, where user's locations setup, ideally equipped with displays, video cameras, depth sensor, microphones and speakers, enables users to communicate and interact remotely experiencing the benefits of a face-to-face meeting in full size. It includes a 3D capture, reconstruction and virtual view synthesys display system.

There are some notable works that realistically exploit the user's appearance for tele-immersion like those developed at UC Berkeley [Kurillo 08] and at GrImage at INRIA [Petit 09]. Both use video cameras array to perform real-time full body 3D reconstructions leading to some weaknesses, like: reconstruction problems due to the lack of accuracy in low-texture or repeated pattern re-

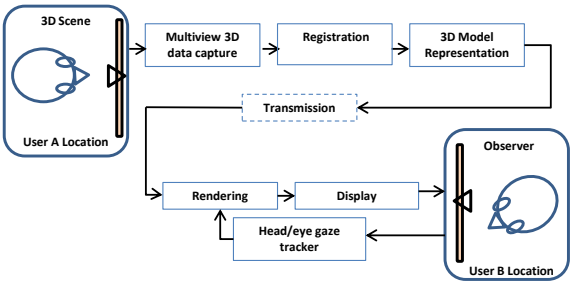


Figure 1. Face to face meeting through technology mediation, line of sight preserving method. Overview of the reconstruction algorithm that aims to continuously generate a realistic body model, transfer the model and reconstruct on a remote common display or virtual environment according, each user's viewpoint by a tracking process. The proposed real-time 3D full reconstruction system combines visual features and shape-based alignment between consecutive point clouds while the mesh model representation is updated incrementally using a new Crust based algorithm.

gions, high cost acquisition data setups, high power computational requirements, and their unsuitability for domestic use. Recent RGB-D reconstruction related works are using alignment and integration approaches based on SLAM sparse methods [Beck 13][Almeida 13]. Henry et al. [Henry 12] combine visual feature matching with ICP-based pose estimation to build a pose-graph which they optimize to create a globally consistent map. Newcombe et al. [Newcombe 11] presented an improved accurate solution known as KinectFusion which uses a new algorithm for real-time dense 3D mapping. KinectFusion integrates depth maps from the Kinect into a "truncated signed distance function" (TSDF) representation. The required alignment to fuse the depth maps is based on the iterative closest point algorithm (ICP), that runs on a GPU for obtaining real time performance.

Our contribution is a real-time 3D full reconstruction system that combines visual features and shape-based alignment between consecutive point clouds while the mesh model representation is updated incrementally using a new Crust based algorithm.

The paper is organized as follows. Section 2 describes the proposed reconstruction methodology, section 3 presents some experimental results and discussion and, section 4 presents the future work and conclusions.

2 MESH GENERATION

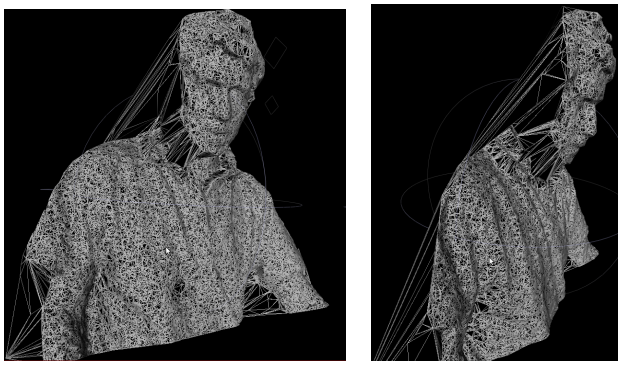


Figure 2. Mesh model using Crust triangulation

An incremental adaptation of Crust algorithm is proposed and enables the addition of new 3D points without having to recompute previous generated meshes. The stitching process relies on integrating new mesh poles as new vertices, on triangulation step and compute triangles only where both surfaces share vertices.

Given a set of registered points $X \in R^3$ sampled from an object surface S, it is possible to approximate its shape by a triangle mesh. The approach, based on a modified Crust algorithm [Amenta 98], uses a set of points P from the medial axis (polos) to extract a subset from the Delaunay triangulation of X that approximate S. The polo points, obtained from the Voronoi vertex or triangles average outer normal's, are positive (p^+) if they lie on the convex side of the surface and negative (p^-) otherwise. Once computed the Delaunay triangulation of $X \cup P$, the surface mesh is estimated by extracting the set of simplices whose vertices belong to X. The proposed approach adds an incremental characteristic to the Crust algorithm as it is efficient viable to add new vertices to a Delaunay triangulation.

Assuming that a set of points X_t were already processed by the Crust algorithm, the set of poles P_t and the Delaunay triangulation are also available [Almeida 11]. To add a new set of sample points X_{t+1} to the surface mesh, avoiding a complete mesh recalculation, the following steps are performed:

Algorithm 1 Crust incremental algorithm

1: P_{t+1}=poles of X_{t+1}
2: Add $P_{t+1} \cup X_{t+1}$ as new Delaunay triangulation vertices
3: Extract triangles whose vertices belong to $X_t \cup X_{t+1}$

The procedure can be applied repeatedly to accommodate any number of point sets X_i. Nevertheless to avoid progressive grow in the number of mesh vertices, points closest to the mesh vertex (i.e. under a given Euclidean distance threshold) are deleted from the input point cloud before the incremental Crust step. Figure 2 illustrates a mesh model using the Crust approach.

Multiview 3D Scan: Recent depth sensor devices, like XBOX Kinect provide 3D measurements and also RGB data, enabling the use of 2D image algorithms. It is possible to improve the 2D feature mapping between consecutive RGB images, associating the respective depth data and creating a 3D feature tracking. The Xbox 360® Kinect™ Sensor combines a RGB camera and a structured light 3D scanner, consisting of an infrared camera and an infrared (IR) laser projector. The depth measurement principle is based on a triangulation process [Freedman 10].

Registration:

The registration process enables to align several 3D point clouds into one same referential to create a global model (Figure 3(b)). To register new 3D point clouds, acquired from different point of views, we perform algorithm 2 steps (Figure 3(a)):

Algorithm 2 Registration algorithm

1: Select one 3D point cloud shape to be the approximate 3D mean shape (ex: scan 0).
2: Align the 3D point cloud shapes:
- Compute the centroid of each 3D point cloud shape (or set of invariant features).
- Align all shapes centroid to the origin.
- Normalize each shapes centroid size.
- Compute the rigid-body transformation using expression (5) to (7) to obtain the rotation R and translation t which best aligns both 3D shapes.
3: Apply the calculated transformation to obtain new approximate 3D mean shape

Considerer the existence of two corresponding 3D points

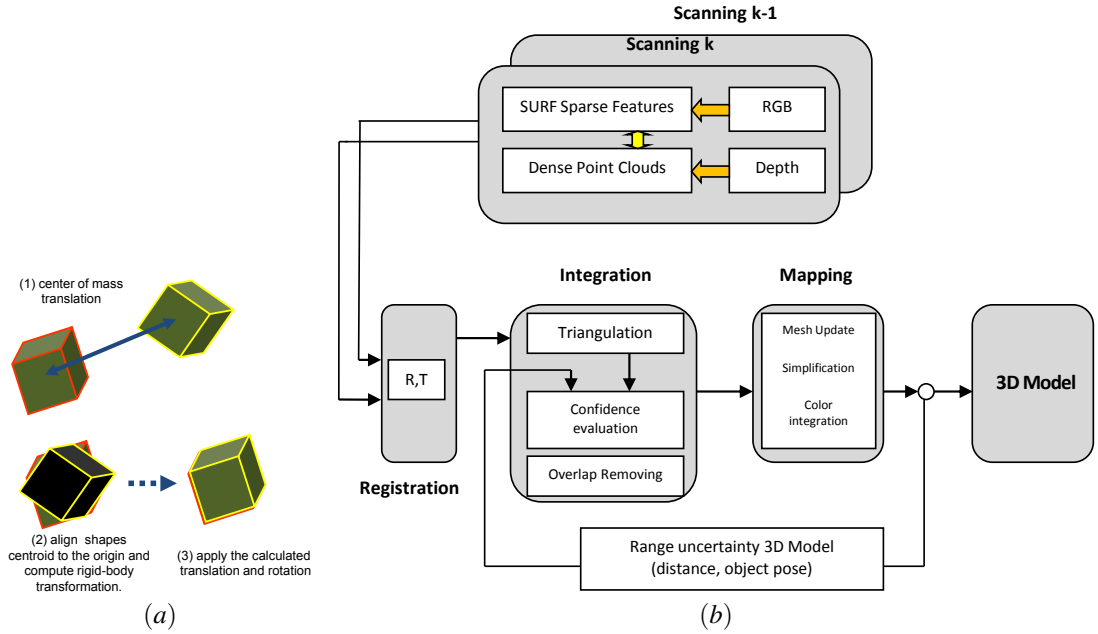


Figure 3. (a) Registration simplified flow. (b) Algorithm overview modules

sets $\{\mathbf{x_i^t}\}$ and $\{\mathbf{x_i^{t+1}}\}$, $i = 1..N$, from consecutive t and $t+1$ scans, which relationship is given by equation (1):

$$\mathbf{x_i^{t+1}} = \mathbf{R}\mathbf{x_i^t} + \mathbf{t} + \mathbf{v_i} \quad (1) \qquad \varepsilon^2 = \sum_{i=1}^{N} \left\| \mathbf{x_i^{t+1}} - \mathbf{R}\mathbf{x_i^t} - \mathbf{t} \right\|^2 \quad (2)$$

R represents a standard 3x3 rotation matrix, **t** stands for a 3D translation vector, and $\mathbf{v_i}$ is a noise vector. The optimal transformation R and t that maps the set $\{\mathbf{x_i^t}\}$ on to $\{\mathbf{x_i^{t+1}}\}$ can be obtained through the minimization of the equation (2) using a least square criterion. The least square solution is the optimal transformation only if a correct correspondence between 3D point sets is guaranteed. Complementary methods are used to robust the correspondence (e.g. RANSAC). The singular value decomposition (SVD) of a matrix can be used to minimize Eq. (2) and obtain the rotation (standard orthonormal 3x3 matrix) and the translation (3D vector) [Arun 87][Challis 95][Eggert 97]. In order to calculate rotation first, the least square solution requires that $\{\mathbf{x_i^t}\}$ and $\{\mathbf{x_i^{t+1}}\}$ point sets share a common centroid. With this constraint a new of equation can be written using the following definitions:

$$\overline{\mathbf{x_i^t}} = \frac{1}{N}\sum_{i=0}^{N} \mathbf{x_i^t} \qquad \overline{\mathbf{x_i^{t+1}}} = \frac{1}{N}\sum_{i=0}^{N} \mathbf{x_i^{t+1}} \quad (3)$$

$$\mathbf{x_{ci}^t} = \mathbf{x_i^t} - \overline{\mathbf{x_i^t}} \qquad \mathbf{x_{ci}^{t+1}} = \mathbf{x_i^{t+1}} - \overline{\mathbf{x_i^{t+1}}} \quad (4)$$

$$\varepsilon^2 = \sum_{i=1}^{N} \left\| \mathbf{x_{ci}^{t+1}} - \mathbf{R}\mathbf{x_{ci}^t} \right\|^2 \quad (5)$$

Maximizing $Trace(\mathbf{R}\,\mathbf{H})$ enable us to minimize the generated equation (5), with **H** being a 3x3 correlation matrix defined by $\mathbf{H} = \mathbf{x_{ci}^{t+1}}(\mathbf{x_{ci}^t})^{\mathbf{T}}$. Considering that the singular value decomposition of **H** results on $\mathbf{H=UDV}^T$, then the optimal rotation matrix, **R**, that maximizes the referred trace is $\mathbf{R= U}$ diag(1; 1; det($\mathbf{UV}^T$)) $\mathbf{V}^T$:

$$\mathbf{R} = \mathbf{U}\mathbf{V}^{\mathbf{T}} \quad (6)$$

The best translation that aligns $\{\mathbf{x_i^{t+1}}\}$ centroid with the rotated $\{\mathbf{x_i^t}\}$ centroid is

$$\mathbf{t} = \overline{\mathbf{x_i^{t+1}}} - \mathbf{R}\overline{\mathbf{x_i^t}} \quad (7)$$

Model Mapping

Suppose that the mapping from the world coordinates to one of the scans of the sequence, is known (ex: scan 0) and it is represented by the transformation $^{\mathbf{0}}\mathbf{T_w}$. As described before, for any consecutive pair of scans (t, t+1) from tracked points it is possible to estimate rotation and translation and combine them into a single homogeneous matrix 4x4, $^{\mathbf{t+1}}\mathbf{T_t}$, $\mathbf{T} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$.

Therefore it is possible to compute equation (8):

$$^{\mathbf{i}}\mathbf{T_0} = {}^{\mathbf{i}}\mathbf{T_{i-1}}{}^{\mathbf{i-1}}\mathbf{T_{i-2}} \ldots .{}^{\mathbf{1}}\mathbf{T_0} \qquad {}^{\mathbf{i}}\mathbf{T_w} = {}^{\mathbf{i}}\mathbf{T_0}{}^{\mathbf{0}}\mathbf{T_w} \quad (8)$$

To update the reconstructed model, each acquired 3D point set is transformed to the world coordinate system using $^{\mathbf{i}}\mathbf{T_w}$. This alignment step adds a new scan to the dense 3D model. Alignment between successive frames enables to track the body position over small displacements.

Correspondence: the described 3D registration method requires the knowledge of point correspondences between the existing 3D points set and the newly acquired point

set. To solve this correspondence problem, we take advantage of the fact that RGB-D sensor provides simultaneously scene 3D information and respective 2D image. We propose the use of "Robust Image Features" (like Bay's Speeded Up Robust Features (SURF) [Bay 06]), which enables the identification of one same point in consecutive images. The association of a visual feature with its 3D point, enables to establish a match between consecutive 3D point clouds.

Although the SURF features enable the establishment of correspondences between points from both sets, illumination and viewpoints changes, together with sensor noise, among others, induce variations on those extracted features that may contribute to errors in the pairing process. This may indeed destroy the transformation estimation process by introducing unacceptable error or leading to no solutions.

For this reason we use the RANSAC algorithm [Fischler 81] to remove false correspondent point pairs that wrongly biases the rigid body transformation estimation. The approach randomly samples three 3D points correspondent pairs from consecutive scans and iteratively estimates the rigid body transformation [Arun 87] until find enough consensus or reach a maximum number of iteration based on the probability of outliers.

The registration method with outliers removal is described in following algorithm 3.

Integration: A new 3D mesh acquired a from different point of view and registered into a 3D global model can lead to two situations: (1) some *non-overlapped* triangles contains new information for the 3D model and (2) some *overlapped triangles* might contain redundant data, or more confident data useful for the model refining. To choose which information is relevant, we evaluate the data based on the uncertainty of range sensor. Sensor accuracy measures are dependent on the incident angle between the measuring ray and the surface distance.

Overlapping segmentation, front face checking and matching: the overlapping region is determined by projecting the pre-built mesh vertices's into the sensor 2D plane, once transformed for the referential of the newly scanned vertices and by checked the intersection area. We could simply re-triangulate all the points on the overlapping region, but due misalignment errors it can result on a bumpy surface. To tackle this challenge we propose an approach, where the triangulations update only happens if it contributes to improve the global model. The process consist in detecting overlapping triangles on the previous scanned range data image and the newly scanned range, and then keep those that provide more information for the model. We associate to each triangle a confidence value based on the measure uncertainty of its 3D vertices. The distance from where sensor acquires the data and the angle from it stands in front a surface are inversely proportional to the confidence (eq. 9):

Algorithm 3 Registration algorithm with outliers removal

Input :X_p, X_q
{assumed correspondent 3D point pairs}
Output :$[R,t]$
{rigid body transformation estimation}
while $(i < MAXITER)$ **do**
  randomly select 3 pairs of points
  $[R_i,t_i] \leftarrow$ estimate 6DOF rigid body transformation for these 3 pairs
  $X'_q = R_i * X_q + t_i$
  {apply the transformation to X_q scan to map it into X_p reference frame}
  $inliers_i = |(X'_q - X_p) < \tau|, number\_of\_inliers_i$
  {determine the set of data points which are within a Euclidean distance threshold τ}
  if $(sizeof(inliers_i) > T_{threshold})$ **then**
    $[R,t] \leftarrow$ re-estimate the transformation model using all $inliers_i$
    $EXIT$
  end if
  if $(number\_of\_inliers_i > bestscore)$ **then**
    $bestscore \leftarrow number\_of\_nliers_i$
    $best\_inliers \leftarrow inliers_i$
    {store cardinality of $inliers_i$ and $inliers_i$}
    update $MAXITER$
  end if
  $i = i + 1$
end while
$[\mathbf{R}, \mathbf{t}] \leftarrow$ re-estimate the transformation model using all points from $best\_inliers$

$$C_i = |\frac{1}{L\theta}| \quad (9)$$

where L is the distance between a 3D point and the range sensor's optical center and θ represents the sensor's pose angle in relation to the surface.

The angle θ is given by equation (10)

$$\theta = arcos(\overrightarrow{n_i}, \overrightarrow{r_i}) \quad (10)$$

where $\overrightarrow{n_i}$ is the normal of a triangle and $\overrightarrow{r_i}$ is the normalized measurement ray from the sensor's optical center to the point.

The confidence measures capture the fact that points close to the sensor, as surfaces close to a fronto-parallel orientation, are typically captured more accurately by range sensors. The normal vector of a point consists of averaging normal vector of triangles formed with pairs of neighbors, and for each new scanned 3D mesh, a list of triangles (3D faces) is be tagged with confidence information related with its 3D point positions. Integration of new triangles will occur, only if, its confidence contributes to improve the 3D model.

Figure 4 depicts the principle of a range sensor, composed by 3 ray measure beams, scanning an object from different positions (2D example). In this case, the range sensor acquires data from 4 different point of views, S_0, S_1, S_2, S_3. For example, due overlapping data measures, between S_0, S_1, S_3 we can incrementally update the global model with the more confident edges (ex: P_{30}, P_{31}, P_{32}).

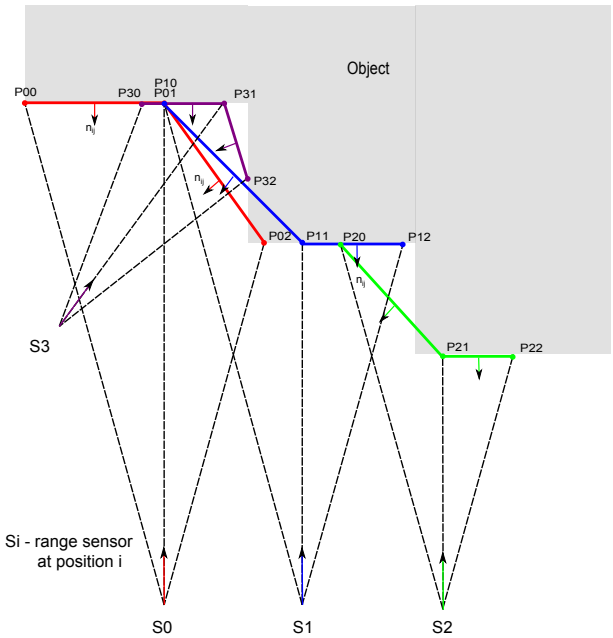


Figure 4. Range sensor, composed by 3 ray measure beams, scans an object from different positions (2D example)

Filtering Methods: depth maps containing holes, inconsistent data in the depth image object boundaries and vibrating behavior at the depth pixel level should be addressed to improve 3D reconstructions. Temporal filtering methods based on time data averaging clearly improves the depth maps quality, although are impractical on real-time applications or where moving objects exist. Several noise removal methods are possible to enhance the Kinect depth maps quality [Tomasi 98][Paris 06], like median filter, bilateral filter, joint bilateral filter, non-local means filter or moving square fitting. For example, the bilateral filter is a non-linear filter based on Gaussian distribution, which reduces the noise smoothing the signal while preserving the edges, however it has a high computational cost.

3 RESULTS

The integration and mesh refining algorithm were previously tested in matlab with noise free point data set and provided useful hints to understand the system. Figure 5 depicts a 3D mesh model of an object (light blue) for which the face triangles normals were computed (red arrows). These triangles and vertices's are projected into the RGB-D sensor plane, here represented by the light green square. The coordinate referential is composed by the blue axes and its intersection is the projection center (referential origin). The face triangles projections are represented in yellow. In Figure 6, the object is rotated slightly around its axes, here represented by light green color. Knowing the rigid transformation, the visible vertices are transformed to match with the previous model and reprojected into the sensor plane, Figure 7. The re-projection of the mesh into image sensor plane enables to detect the triangle intersection and preserve triangles with higher confidence.

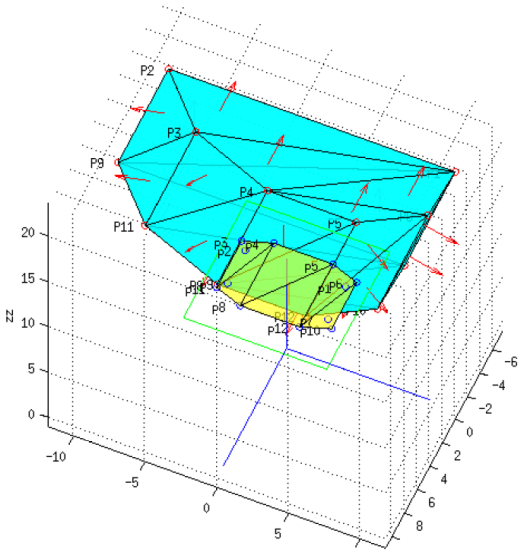


Figure 5. Fixed range sensor scanning an object

In Figure 9 we show an example of correspondence between consecutive image features using SURF method (white lines indicate correspondent point).

Figure 10 depicts a sequence of scans that creates a 3D per-

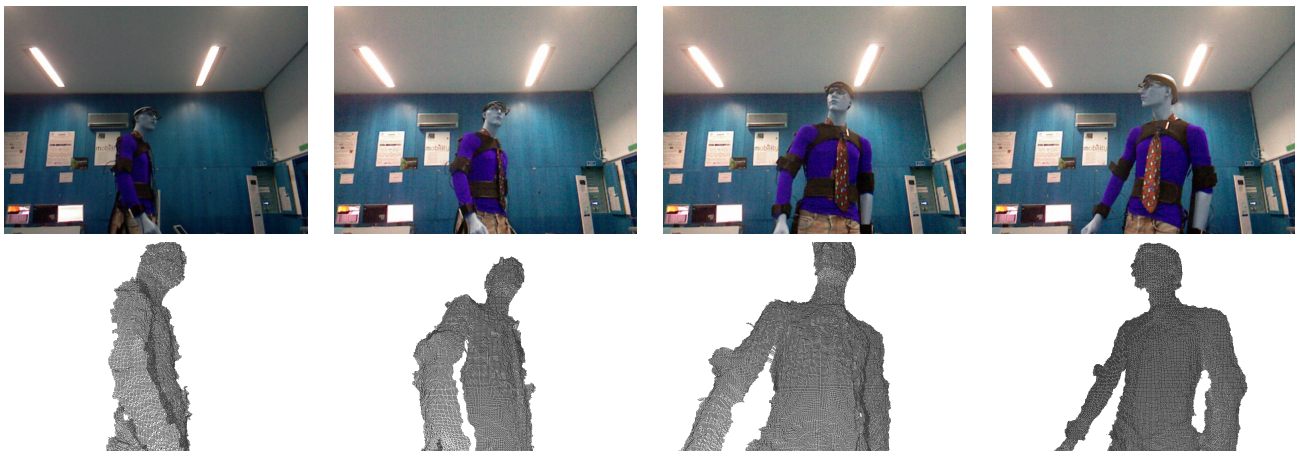


Figure 10. Sequence of mesh models to be integrated, triangulation based on depth data sensor grid structure and depth information.

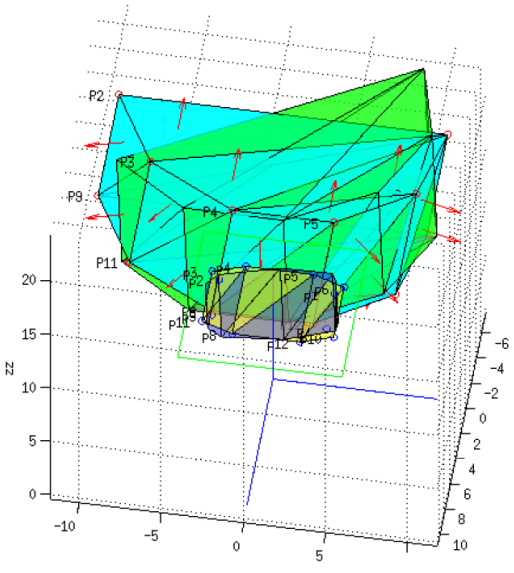


Figure 6. Moving object

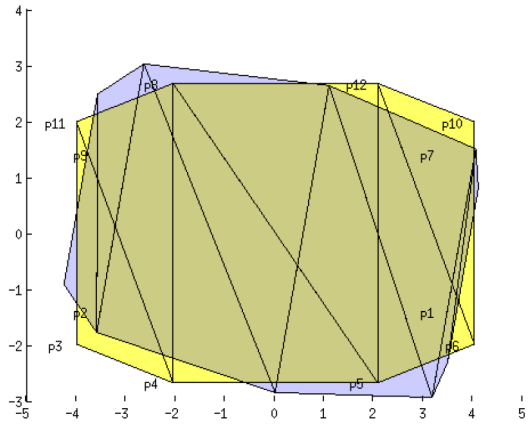
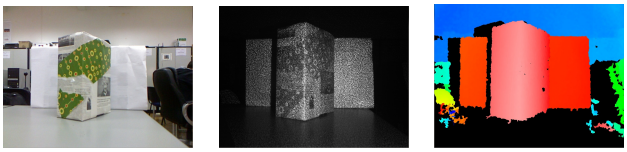


Figure 7. Mesh re-projection into image sensor plane to detect triangle intersection. Preserve triangles with higher confidence.



(*a*) (*b*) (*c*)

Figure 8. (a) RGB image. (b) IR monochromatic image with speckles pattern projected onto a scene. (c) Depth map with distances associated to colors.

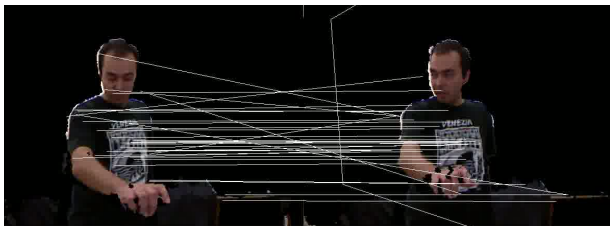


Figure 9. SURF features matched on consecutive time frames

son model. On the top row we present RGB images of the scene and in lower row snapshots of the respective meshes, generated in real time. The mesh triangulation is based on depth data sensor grid structure and depth information. To achieve the real time characteristic, we programmed in OpenGL for Embedded Systems (OpenGL ES) as it enables vertex buffers to be processed in parallel as a single entity. GPU shaders and OpenCV [OpenCV 15] were also used.

Figure 11 shows a reconstructed 3D model. It results from several 3D point clouds fused in real time after applying successive 3D rigid body transformations, mesh refining integration and rendering.

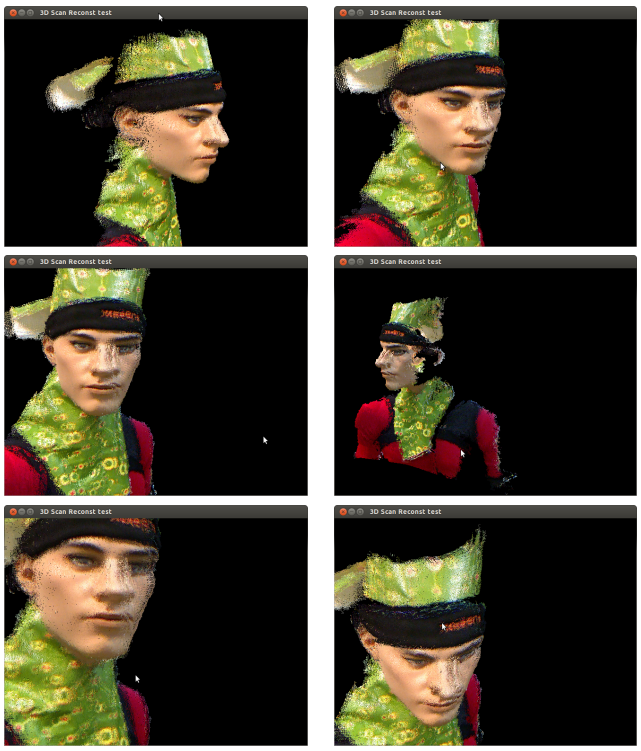


Figure 11. Synthesized views of a on-line 3D reconstructed model dependent of observer point of view.

3.1 Discussion

Processing real data allowed us to identify some noise sources that can affect the algorithm. For example, SURF points can generate erroneous matches due image noise and they are more common on body boundaries (Figure 9 presents some wrong diagonal links for an almost pure vertical axis body rotation). The body to be reconstructed should be segmented from background static areas using a motion filter. Scale-invariant feature transform (SIFT) [Lowe 04] was also tested and presented better accuracy as key feature descriptor, although we have chosen SURF method in order to achieve the real-time characteristic. The kinect system imaging geometry introduces structural errors that are function of the distance to the object and the sensor orientations relative to the object surface. A proper calibration of the RGB-D sensor is essential to improve results. Stereo calibration procedures were used to estimate the intrinsic parameters of both RGB and IR (depth) cameras, as the relative transformation (R,T) between them. The estimated camera's parameters and transformations enabled us to align Kinect™ both RGB with IR (depth) cameras and obtain more reliable data information (as depicted in Figure 8). The proposed reconstructed 3D model approach enables to generate any virtual synthesized view for an observer that moves in front of a display, that is, a required augmented reality (AR) functionality.

4 CONCLUSION

A free viewpoint system framework is proposed to generate view dependent synthesis based on scene 3D mesh model. Our approach explores virtual view synthesis through motion body estimation and hybrid sensors composed by video cameras and a low cost depth camera based on structured-light. The solution addresses the geometry reconstruction challenge from traditional video cameras array, that is, the lack of accuracy in low-texture or repeated pattern region. We present a full 3D body reconstruction system that combines visual features and shape-based alignment. Modeling is based on meshes computed from dense depth maps in order lower the data to be processed and create a 3D mesh representation that is independent of view-point. Research contributions include a new incremental version of Crust algorithm that efficiently adds new vertices to an already existing surface without having to recompute previous generated meshes and a topological incremental reconstruction approach based on confidence measures that avoids redundant data information computation.

With this on-line reconstructed 3D model, we can provide synchronous point of view for an observer that moves in front of a display of a face-to-face meeting application, thus enhancing the presence sensation. Future work includes framework usability tests for a telepresence meeting application. This work presents an on-line incremental 3D reconstruction framework that can be used on low cost telepresence applications, augmented reality (AR) or human robot interaction applications.

References

[Almeida 11] Luis Almeida, Filipe Vasconcelos, João Barreto, Paulo Menezes, and Jorge Dias.

On-line incremental 3d human body reconstruction for hmi or ar applications. In *CLAWAR 2011: 14th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machine*. Paris, France, September 2011.

[Almeida 13] Luis Almeida, Paulo Menezes, and Jorge Dias. *Handbook of Research on ICTs for Human-Centered Healthcare and Social Care Services*, chapter Augmented Reality Framework for the Socialization between Elderly People, pages 430–448. IGI Global, 2013.

[Amenta 98] Nina Amenta, Marshall Bern, and Manolis Kamvysselis. A new voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 415–421. ACM, New York, NY, USA, 1998.

[Arun 87] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, September 1987.

[Bay 06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *In ECCV*, pages 404–417. 2006.

[Beck 13] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013.

[Challis 95] J. Challis. A procedure for determining rigid body transformation parameters. *Journal of Biomechanics*, 28(6):733–737, jun 1995.

[Eggert 97] D. W. Eggert, A. Lorusso, and R. B. Fisher. Estimating 3D rigid body transformations: a comparison of four major algorithms. *MAchine Vision and Applications*, 9:272–290, 1997.

[Fischler 81] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[Freedman 10] Barak Freedman, Alexander Shpunt, Meir Machline, and Yoel Arieli. Depth mapping using projected patterns, May 2010.

[Henry 12] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *I. J. Robotic Res.*, 31(5):647–663, 2012.

[Kurillo 08] G. Kurillo, R. Vasudevan, E. Lobaton, and R. Bajcsy. A framework for collaborative real-time 3d teleimmersion in a geographically distributed environment. In *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, pages 111 – 118. dec. 2008.

[Lowe 04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60:91–110, November 2004.

[Newcombe 11] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 127–136, Washington, DC, USA, 2011. IEEE Computer Society.

[OpenCV 15] OpenCV. http://opencv.org, 2015.

[Paris 06] Sylvain Paris and Frédo Durand. A fast approximation of the bilateral filter using a signal processing approach. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part IV*, ECCV'06, pages 568–580, Berlin, Heidelberg, 2006. Springer-Verlag.

[Petit 09] Benjamin Petit, Jean-Denis Lesage, Clément Menier, Jérémie Allard, Jean-Sébastien Franco, Bruno Raffin, Edmond Boyer, and François Faure. Multicamera real-time 3d modeling for telepresence and remote collaboration. *INTERNATIONAL JOURNAL OF DIGITAL MULTIMEDIA BROADCASTING*, 2010:247108–12, 2009.

[Tomasi 98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846, Jan 1998.